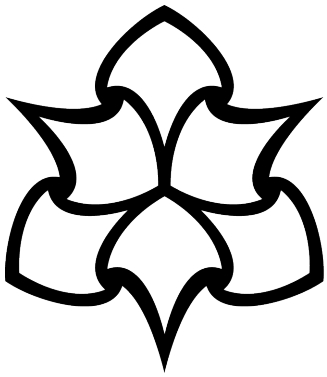

Information Environment Formative Evaluation

EDNER+

Quality Study

Comparison of the RDN Hubs and Google as search tools

EDNER+ Deliverable

August 2004

Contact

Professor Peter Brophy

CERLIM

p.brophy@mmu.ac.uk

Quality Study

Version control

Version	*Date of release*	*Status/Notes*
1.0	27 July 2004	Draft
2.0	August 2004	Minor revisions in line with feedback from JISC

This document should be cited as:

Brophy P, and Markland M, (2004) EDNER+: Information Environment Formative Evaluation: Quality Study 1. Comparison of the RDN Hubs and Google as search tools Deliverable Q1: Manchester: CERLIM (Centre for Research in Library & Information Management)

Information Environment Formative Evaluation EDNER+ is a one year project being undertaken by the Centre for Research in Library & Information Management (CERLIM) at the Manchester Metropolitan University and the Centre for Studies in Advanced Learning Technology (CSALT) at Lancaster University. Details of the project's work and copies of published reports are available at http://www.cerlim.ac.uk/projects/iee/index.php

Contents

Quality Study: Comparison of the RDN Hubs and Google as search tools

1. Aim and objectives

The **aim** of this study is to make preliminary comparisons of the the RDN Hubs and Google as search tools within an academic context.

The **objectives** were:-

- To ask Hubs Managers to supply the five words or phrases most commonly used to search their Hubs.

- Using this real data, to carry out these five most popular searches on each RDN Hub and note the first five search results displayed.

- To carry out an identical set of searches on Google, and note the results obtained on the first three pages of results (i.e. 30 results)

- To clarify whereabouts, if at all, the Hubs' results appear in the Google results.

- To clarify whether any of the RDN results are unobtainable through Google.

2. Rationale for the study

As is shown in the following brief review of the literature, the student's preference for search engines (and particularly Google) as an information seeking tool has been well documented, both within the EDNER project and elsewhere. Also, EDNER has found that awareness and use of the RDN Hubs within the academic community is generally low. Given the RDN's claim that *'In contrast to search engines, the RDN gathers resources which are carefully selected, indexed and described by specialists in our partner institutions'* and that therefore *'You can be confident that your search results and browsing will connect you to Web sites relevant to learning, teaching and research'*, (RDN 2004) this raises the important question of the degree to which

students are missing quality academic resources because of their preference for search engines. Do they need a tailored service which delivers only quality information to them, or does Google, a popular search engine among students, in fact deliver the same, similar or even more appropriate kinds of resources?

The quality of resources used by students, teachers and researchers relates to a key theme of the EDNER+ project, that of aiming to examine the achievements of the JISC Information Environment (IE) in securing and promoting high quality information content and services. As a first step towards investigating this complex issue it was decided to conduct this small and very focused study to see if it would reveal any answers, or if not, would clarify what further questions could usefully be asked. It was not intended to make any premature judgments about the value of the RDN Hubs or Google, simply to compare how each responded to a search query. Furthermore it was not intended that we would make value judgements about the quality of the resources displayed.

3. Literature Review

Students like search engines. There really is no doubt about this assertion. Brophy et al. (2001), for example, report an EDNER user testing study where university students were set fifteen online information seeking tasks, but given no guidance on how to go about finding answers. They report that the majority of students went first to a search engine to help them find the information they needed. In fact over 70% of their users regularly turned to search engines first to help them find information. Google was way ahead of the others (45%), with its nearest rival Yahoo at only 9%. This finding is reiterated by Banwell et al. (2004) in the JUSTEIS project. They state that *'search engine use predominates over all other EIS; …. Over the three cycles there has been a shift in the preferred search engine from Yahoo to Google'* (p311). UK research is echoed elsewhere. OCLC, for example (2002 p9) report that 79% of students use search engines for all or most assignments, 50% use web portals, and 40% use 'course specific websites'. Again a predominant use of the search engine as primary information seeking tool is demonstrated.

This predominance of search engine use might not be so important if students were also using subject gateways as an information seeking tool, but both UK projects

suggest that this is not the case. Banwell et al for example found that *'Use of JISC Gateways, such as SOSIG and OMNI, is very limited* (ibid p310). Furthermore Jones and Goodyear (2003) suggest that lecturers too did not use them much, and are therefore, by inference, unlikely to encourage their students to use them. They say *'Gateways were little used or known about. Three interviewees knew of the subject gateway relevant to their discipline area. Of these two made little use of the relevant gateway (SOSIG, HUMBUL), whilst the third stated an explicit preference for Google over the gateway (BIOME)'*(p 11) They also uncovered a lack of awareness among some subject librarians.

Another factor in information seeking behaviour which is particularly relevant to this study, is that students rarely look beyond the first few pages (or even the first page) of search engine results. Craven and Griffiths (2002 p183) for example report two independent studies with two very different groups of students (one visually impaired, one not) both of which report that the average number of pages evaluated was only one. They summarise that *'Overwhelming numbers of both sets of participants looked at only the first page of results'* while the Craven study goes even further to report that *'All looked at only the first page of returned results – usually just the first two or three hits, before pursuing one of these links further or reformulating the query'* a pattern of behaviour which, they say, is reiterated in other studies. Although not reporting this precise finding, Spink et. al. (2002, p 108) illustrate that the incidence of 'one page only' viewing is certainly a rising trend. They give figures from three iterations of an extensive longitudinal study which show clearly that between 1997 and 2001 the number of users viewing only the first page rose from 28.6% to 50.5%, while the number viewing three or more pages fell from 51.9% to 29.2%, with over 70% of users looking at 'two pages or fewer' by 2001 (ibid p107). In a separate 1998 study into search engine use, the same team discovered that 58% of users viewed page 1 only, while only 4% viewed more than 4 pages (Jansen 1998). They pose the interesting questions *'Were they so satisfied with the results that they did not needed to go viewing more? Is the precision that high? And are the users after precision – few answers were good enough? Or did they just give up? Who knows?*' In 2004 we still cannot answer these key questions with any degree of certainty, but we do continue to find that 'satisficing', using something that is good enough rather than seeking out an optimal result, is the norm and particularly so among students. As Brophy et al (2004 p14) report *'Unlike the academic researcher who usually has a requirement to locate <u>the</u> key paper in his or her field in order to ensure that an*

approach or finding has not been overlooked, learners are often satisfied with "any" resource which comes close to meeting their expressed need – and there are often many alternatives available.' This finding might in fact mitigate in favour of resources such as the RDN. If high quality, precisely targeted resources can be delivered to the learner with the minimum of effort on his or her part, this should overcome the two obstacles of reluctance to look beyond the first page or two of results, and making do with whatever they find there. In fact, a service which delivers this sounds like the answer to the student's dream.

With these factors about user behaviour in mind it seems worthwhile conducting a study to see what students are actually presented with when they search for information for their academic studies online, and what comparisons can be made and conclusions drawn about how search results from the RDN Hubs and Google compare.

4. Method

The Managers of the eight RDN Hubs were asked to provide the five words or phrases most commonly used to search their Hub, and all did so. In fact fifty search terms were provided, of which thirty-one were single words and the remainder two or three-word phrases. The first five terms provided were then run as simple searches on both the appropriate Hub and on Google, and the results recorded. Advanced searching was not used. The first five Hub results were compared to the 'hits' on the first three pages of Google results (namely thirty results) to see if the Hub results appeared within these Google hits. Each Hub search therefore produced a set of twenty-five results, which are reported in the following results section. For illustrative purposes, a shortened set of three searches and three results for each search is shown in full. These indicate the kinds of resources found.

When a Hub resource did not appear in the first three pages of Google, a further check was made to see if it could be accessed through Google, or if it was part of the 'hidden web'.

5. Results

5.1 Google

Searching Google

The Google home page offers a simple keyword search box surrounded by a range of different ways in which to refine the search. There is an advanced search facility which permits various Boolean functions, searching by language, file format, numeric range, date and domain. There is a 'safe search' to filter out adult content, a page specific search and a topic specific search. It is possible to search images, shopping sites, to join groups, catch up on the latest news, to set preferences for how results are displayed and to use translation tools. The 'I'm feeling lucky' feature takes the searcher straight to the first web page returned for the query. There are resources for advertisers, web publishers and businesses, and information on how to use Google to search your own home website. All of this complexity is achieved whilst still retaining a clear and simple homepage dominated by a simple search box and a 'search button'.

Google ranks its results by use of its PageRank software. This is an automated process designed, the company says, to 'put the most relevant and reliable results first'. The premise of the process is that the volume of links to a webpage can be taken as an indicator of the page's value. The process is described thus *'In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves 'important' weigh more heavily and help to make other pages 'important'.'* (Google 2004). This is combined with other unspecified but *'sophisticated text-matching'* techniques to ensure that *'all aspects of the page's content (and the content of the pages linking to it)'* are considered when determining *'if it's a good match for your query'*.

Google's explanatory publicity draws upon many terms which are very much human value judgments to describe how it ranks its results. It speaks of importance,

relevance, reliability, high-quality. Yet at the same time Google prides itself in the lack of human involvement in its PageRank system, and the non-manipulation of results, seeing this as a ensuring *'objective information untainted by paid placement'*. The underlying process to all of this, however, does depend upon human involvement and indeed is open to a degree of manipulation, (whether paid or not) because it is based upon the cumulative frequency of links to and from web pages, whose presence is, after all, the result of decisions made by people. Furthermore Google equate 'linking to a page' as 'assigning importance', but 'importance' does not necessarily indicate quality.

With these factors in mind then, it is interesting to observe how the 'quality' search results for Google compare to those of the Hubs which are specifically chosen for the academic quality of their content.

5.2 Searching the RDN

The RDN offers the searcher two choices; a cross-search of all the Hubs from the RDN homepage using the 'Resource Finder' feature, and a search of each of the individual Hubs. Although a comprehensive comparison of search results using these two techniques did not form part of this study, a small sample of Resource Finder searches was carried out, and this is reported in section 5.11.

5.3 ALTIS http://www.altis.ac.uk/

ALTIS is a guide to Internet resources in hospitality, leisure, sport and tourism.

Results of the ALTIS Searches

1 ALTIS result appears on the 1st page of Google results

24 results do not appear in the first three pages of Google results

ALL websites could be accessed from Google

First Three ALTIS Results	Found on Google page			Access from Google?
	1	2	3	
Search term = world cup **ALTIS hits = 10 Google hits = 7,650,000**				
UEFA.com: Europe's football website http://www.uefa.com/	N	N	N	Y
Diplomatic background to the 1996 Football World Cup http://www2.umist.ac.uk/sport/polley182.html	N	N	N	Y
British Biathlon Union (BBU) http://www.britishbiathlon.com/	N	N	N	Y
Search term = world cup 1966 **ALTIS hits = 1 Google hits = 291,000**				
Diplomatic Background to the 1966 Football World Cup http://www2.umist.ac.uk/sport/polley182.html	N	N	N	Y
Search term = football **ALTIS hits = 135 Google hits = 42,600,000**				
Soccer ball world http://www.soccerballworld.com/	N	N	N	Y
Bibliography of Association football in Wales http://www.staff.ucsm.ac.uk/mjohnes/biblio.htm#explan	N	N	N	Y
Action for Blind People : Sport http://www.afbp.org/Sports/	N	N	N	Y

Searching ALTIS

ALTIS offers a simple search on its home page, which has a default setting to 'Any resource type'. There is a second box which permits the searcher to narrow the search to one particular resource category, such as full text journal or organizational website. A browsing option is currently being developed, and there is a 'help with searching' facility. The logic behind how ALTIS presents its results is unclear, but the

list of results is not arranged alphabetically. Each record has a resource title which links to the underlying website, a brief description which includes a note of any special facilities which might be needed in order to view the resources, a list of keywords associated with the resource and a statement of which of the ALTIS resource categories the resource is located in.

Comments

The ordering of results by ALTIS is puzzling. For example in the 'football' search the first hit is a US website containing information on the design, construction etc. of footballs, the second a bibliography of football in Wales, and the third is a website for blind people which gives details of sporting events in which they can take part. Fourth comes UEFA, and FIFA only seventh. The 'world cup' search includes a number of sports, including biathlon and cricket.

The 'world cup' search displays the UEFA web site as the first hit; in fact, of course, this relates only to the European Cup and does not appear to contain a link to the (football) world cup.

This was the only Hub where a search resulted in only one hit, a conference paper written by an academic, and the search was actually a refinement of the previous search. Considering that this is one of the most popular searches on this Hub, this is an example of where Hub managers may wish to put effort into seeking out further resources to offer their users.

Google offers a variety of websites for all of the searches, with archive material of actual events, statistics, histories, comment (some very biased, some more objective), gambling, joke sites, and books and videos for sale. Google covers many sports, football, (American, Australian, African, Asian and Scottish), skiing, rugby, paragliding, and hockey.

5.4 ARTIFACT http://www.artifact.ac.uk/

ARTIFACT is a guide to resources in the Arts and Creative Industries

Results of the ARTIFACT Searches

3 ARTIFACT results appears on the 1st page of Google results

22 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three ARTIFACT Results	Found on Google page			Access from Google?
	1	2	3	
Search term = Manchester **ARTIFACT hits = 42 Google hits = 16,600,000**				
Aliens.au http://www.anat.org.au/pages/archived/1997/aliens/index.html	N	N	N	Y
Association of national dance academies http://www.anda.org.uk/	N	N	N	Y
AT03 Artranspennine http://www.artranspennine.org.uk/index.html	N	N	N	Y
Search term = Photography **ARTIFACT hits = 361 Google hits = 29,600,000**				
19th Century photography of Ancient Greece http://www.getty.edu/research/conducting_research/digitized_collections/garyedwards/	N	N	N	Y
A Grand Design: the art of the Victoria and Albert Museum http://www.mfa.org/exhibitions/london/	N	N	N	Y
A.P.C. http://www.apc.fr/	N	N	N	Y
Search term = Cindy Sherman **ARTIFACT hits = 10 Google hits = 428,000**				
Broad Art Foundation http://www.broadartfoundation.org/	Y			
Cindy Sherman http://www.temple.edu/photo/photographers/cindy/sherman.htm	Y			
Cindy Sherman : from dream girl to nightmare alley http://www.salon.com/media/1997/12/08media.html	Y			

Searching ARTIFACT

ARTIFACT offers a simple search on its home page, a link to an advanced search (by category and/or resource type) and a choice of three options 'any word, part of a word or phrase'. It is also possible to browse by category from the home page. Results are presented as an alphabetical list, and each has a hypertext link to the underlying website, a brief description, a link to a full record, and a flag symbol indicating the nationality of the site. The full record includes extra information such as any special software which is needed to view the website, or its relationship to a larger resource.

Commentary

All of the ARTIFACT 'Manchester' search results relate to Manchester in the Northwest of England, and all are, of course, connected with the Arts. On the face of it, the most puzzling result is the first, Aliens.au, which proves not to be a sci-fi site but an Australian participant in a cultural event held in the Northwest, apparently in 1997 although the site is undated. Most of the Google hits relate to the UK city of Manchester and its institutions; the Universities, local government offices, football clubs etc. There are links to Manchester in Connecticut, and to a Los Angeles hockey club. The closest link to the Arts is the very last one on page 3, which is to an art gallery.

The ARTIFACT 'photography' search produced some links whose relation to the search term appears quite tenuous. There are, for example, links to an African art exhibition, a Paris fashion house and an expressionist painting website. High on the Google 'hits' list are links to major institutions such as the New York Institute of Photography or the American Museum of Photography, to journals, to online galleries and to niche collections, such as the NASA gateway to astronaut photography of the earth, as well as to information about digital cameras.

The Cindy Sherman search is particularly interesting because the 'hits' on ARTIFACT and Google correspond quite closely. This is possibly because the search term is a particular photographer with a fairly uncommon name, so there is less scope for Google to pick up a large number of 'general' broadly based websites than for a

search such as ‘football’. Indeed the Google hits contain many resources which would seem to contain informative and useful materials about this photographer.

5.5 BIOME http://biome.ac.uk/

BIOME offers resources in the Health and Life Sciences

Results of the BIOME Searches

3 BIOME results appear on the 1st page of Google results

1 BIOME result appears on the 2nd page of Google results,

21 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three BIOME Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = AIDS **BIOME hits = 784 Google hits = 24,200,000**				
3TC – overview http://www.aidsmap.com/en/docs/66A23363-D38A-4B17-AB29-72712E33038B.asp	N	Y		
A simple guide to orthopaedics and trauma http://www.worldortho.com/database/intro-orth/main.html	N	N	N	Y
AAD patient information http://www.aad.org/pamphlets/index.html	N	N	N	Y
Search term = diabetes **BIOME hits = 344 Google hits = 14,400,000**				
2Ayurvedic interventions for diabetes mellitus : a systematic review http://www.ahcpr.gov/clinic/epcsums/ayurvsum.htm	N	N	N	Y
A diabetes website http://www.adiabeteswebsite.com/index.htm	N	N	N	Y
AACE guidelines for the management of diabetes mellitis http://www.aace.com/clin/guidelines/diabetes_2002.pdf	N	N	N	Y
Search term = museum **BIOME hits = 912 Google hits = 39,300,000**				
2000 BG Wildlife Photographer of the year http://flood.nhm.ac.uk/cgi-bin/wildwin/2000/htdocs/	N	N	N	Y
24 Hour Museum http://www.24hourmuseum.org.uk/	N	N	N	Y
AAHO-HNS virtual museum http://www.entlink.net/museum/	N	N	N	Y

Searching BIOME

Biome has a simple search box on the home page, which produces results ranked alphabetically by title. There are also links to browsing, to help with searching and to an interesting advanced search facility. This offers the opportunity to search by different resource types, to deselect listed subject databases, and to rank search results by relevance, title or resource type. The default advanced search is 'sort by relevance ranking' which is in contrast to the results for a simple search which are ranked alphabetically. The BIOME records give a full explanation of the website content, and a list of link words and phrases which appears to be to similar sites, which could presumably be found by browsing.

Comment

The AIDS results are interesting. The first is a fact sheet on (HIV) AIDS, the second a guide to audiovisual aids, and the third another fact sheet on (HIV) AIDS and dermatology. We have no way of knowing in what precise context users carried out the search, but the whole results list is a mix of these two interpretations of the search term. In contrast, the first three Google pages are almost entirely made up of resources to do with HIV AIDS, its prevention and treatment, children and AIDS, law and human rights for AIDS sufferers, conferences and memorials. The only exception is the 26th hit, a site offering Bible study aids.

The resources on the BIOME Diabetes list provide largely mainstream medical and nursing information – apart from one veterinary site which reminds us that cats too can be diabetic. Google offers links to a range of national and international organizations worldwide, to academic and scholarly journal articles and to diabetes charities.

The third search term 'museum' produces a wide range of museum collections and exhibitions within the context of the biomedical sciences; everything from herbs and herbal medicine to a virtual museum of the ear nose and throat, from bat biology to the study of mites and ticks. Google, as might be expected offers links to the major museums of the world; the Metropolitan Museum of Art in New York, the Guggenheim, the Natural History Museum – and the intriguing Museum of Hoaxes.

The Google results, therefore, while linking to world renowned institutions lack the context of the BIOME results.

5.6 EEVL http://www.eevl.ac.uk/

EEVL provides a guide to Engineering, Mathematics and Computing resources

Results of the EEVL Searches

3 EEVL results appear on the 1st page of Google results

3 EEVL results appear on the 2nd page of Google results

2 EEVL results appear on the 3rd page of Google results

17 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three EEVL Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = renewable energy **EEVL hits = 486 Google hits = 1,670,000**				
Renewable energy world http://www.jxj.com/magsandj/rew/index.html	Y			
US Department of Energy Office of Energy Efficiency and Renewable Energy http://www.eere.energy.gov/	Y			
World-wide Information System for Renewable Energy http://wire0.ises.org/	N	N	N	Y
Search term = cable stayed bridge **EEVL hits = 25 Google hits = 84,500**				
Honshu-Shikoku Bridge Authority http://www.hsba.go.jp/e-index.htm	N	N	Y	Y
LUSAS Bridge Case Study Index http://www.lusas.com/case/bridge/index.html	N	N	N	Y
LUSAS Contents Page http://www.lusas.com/toc.html	N	N	N	Y
Search term = fatigue testing **EEVL hits = 252 702,000**				
Statistical Loads Data for Cessna 172 Aircraft using the Aircraft Cumulative Fatigue System (A Report) http://aar400.tc.faa.gov/aar-430/reports/01-44.pdf	N	N	N	Y
The Effect of a Hard Coating on the Damping and Fatigue Life of Titanium *https://research.maxwell.af.mil/papers/ay2003/afit/AFIT-GAE-ENY-03-12*	N	N	N	Y
Poulter laboratory : Aircraft Safety Research http://www.cp.umist.ac.uk/	N	N	N	Y

Searching EEVL

The EEVL homepage offers a search box defaulted to ‘all’ but with the option to change to ‘any’ or ‘phrase’, and to display either a full record, or a record title only. There are options to search the whole EEVL catalogue, or to refine the search to ‘key sites’ of ‘particular importance’ or ‘outstanding quality’, or to widen it to ‘search the sites within the EEVL catalogue’. This provides flexible and sophisticated options for devising a search strategy. EEVL also offers many other features. Hits are ranked by relevance, and each shows a substantial record which can be further extended to reveal other metadata such as country of origin, language, resource type etc.

Comments

The EEVL ‘renewable energy’ search again showed close correspondence with the Google search. The Google pages for this search term have many links to Government websites in the USA, UK, EU and Australia, and to companies offering products and services.

The EEVL ‘cable stayed bridge’ search results are quite specialized, and mostly case studies from a particular company offering bridge design and assessment software. Google offers more general information on this and other types of bridge from a variety of companies and other information resources.

The ‘fatigue testing’ EEVL search produced research papers and case study reports, whereas again Google offered more general information mostly from companies but also from some university research centres.

5.7 GEsource http://www.gesource.ac.uk/home.html

GEsource is resource for Geography and the Environment

Results of the GEsource Searches

2 GEsource results appear on the 1st page of Google results

3 GEsource results appear on the 2nd page of Google results

2 GEsource results appear on the 3rd page of Google results

18 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three GEsource Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = migration GEsource hits = 425 Google hits = 8,500,000				
Agricultural Research in West Africa http://www.rzuser.uni-heidelberg.de/~udemirag/ (File not found on 23.07.04)	N	N	N	
Are Immigrants Leaving California? http://www.urban.org/UploadedPDF/are_immigrants_leaving_ca.pdf	N	N	N	Y
Asian Research Center for Migration (ARCM) http://www.chula.ac.th/INSTITUTE/arcm/main.htm	N	N	N	Y
Search term = desertification **GEsource hits = 114 Google hits = 385,000**				
Aral Sea Homepage http://www.dfd.dlr.de/app/land/aralsee/	N	N	N	Y
Aral Sea Studies http://www.ce.utexas.edu/prof/mckinney/papers/aral/aralhome.html	N	N	N	Y
Arid Lands Newsletter http://ag.arizona.edu/OALS/ALN/ALNHome.html	N	Y		
Search term = global warming **GEsource hits = 436 Google hits = 2,030,000**				
American Forests http://www.americanforests.org/	N	N	N	Y
Arctic Climate System Study http://acsys.npolar.no/	N	N	N	Y
Biomes 1 : Lecture 10 : http://www.valdosta.edu/~grissino/geog4900/lect_10.htm	N	N	N	Y

Searching GEsource

The GEsource home page has links to five ‘gateways’ next to a simple search box. This is supplemented further down the page by various refining tools, such as the selection of subject area and ‘gateway’, and also an advanced search which further enhances this by the addition of a choice of search fields (title, description, keywords). There is a further option to carry out a spatial search by map location, by geographic name or by coordinates. Results are presented in alphabetical order.

Comments

The results for the GEsource and Google migration searches were broadly of like kind, with Google offering information on the migration of people, animals, birds and insects, though from different organizations to those found on the GEsource site.

The desertification results were rather different with GEsource tending towards research centres and scholarly articles, while Google offered a very international mix of governments, universities and other bodies such as Greenpeace.

The ‘global warming’ searches on GEsource offer very specific and focused information from particular scientific or academic bodies, with the third apparently being lecture notes, last updated four years ago, from the little-known Valdosta State University. Google again offers government, and organizational websites, and a number of press and newspaper articles.

5.8 HUMBUL http://www.humbul.ac.uk/

HUMBUL offers resources in a range of Humanities subjects including among others, Archaeology, Classics, History, Linguistics, Philosophy and various language studies.

Results of the HUMBUL searches

4 HUMBUL results appear on the 1st page of Google results

1 HUMBUL result appears on the 2nd page of Google results

1 HUMBUL result appears on the 3rd page of Google results

19 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three HUMBUL Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = Shakespeare **HUMBUL hits = 102 Google hits = 6,750,000**				
All Shakespeare http://www.allshakespeare.com/	Y			
American and English Literature internet resources http://library.scsu.ctstateu.edu/litbib.html	N	N	N	Y
AS guru : English http://www.bbc.co.uk/education/asguru/english/	N	N	N	Y
Search term = medieval **HUMBUL hits = 501 Google hits = 6,940,000**				
Osterreichisches Staatsarchiv http://www.oesta.gv.at/	N	N	N	Y
The disavowal of Jewish identification in 'Piers Plowman' B Text http://web.english.ufl.edu/exemplaria/preprints/goldstein/goldfra.htm	N	N	N	Y
AARHMS Archive http://www.uca.edu/divisions/academic/history/aarhms/default.html	N	N	N	Y
Search term = poetry **Humbul hits = 469 Google hits = 29,200,000**				
Copyright http://www.udc.es/dep/lx/cac/sopirrait/	N	N	N	Y
Abel Martin: revista de estudios sobre Antonio Machado http://www.abelmartin.com/	N	N	N	Y
The Academy of American Poets http://www.poets.org/	Y			

Searching HUMBUL

From the homepage HUMBUL offers simple searching which can be refined by the selection of 'all', 'any', or 'phrase' searches, and a browsing option. Results are returned as a list of basic records containing the title, location, and description of the resources. At the end of the description there is a 'View Full Record' link which links to the full record for the resource. Once a search has been carried out the results can be further refined by resource type, subject, language or period. Results are presented in alphabetical order.

Comments

The Humbul 'Shakespeare' search offers links to large resources, some of which claim to be 'portals', and to broad, general Shakespeare information sites. There are sites linked to BBC broadcasts on Shakespeare, and an interesting link to a BBC educational resource targeted very precisely at AS and A Level students. The Google search results are very wide-ranging and also offer much general information on the works of Shakespeare, the history of his life and times, theatres, and again links to a number of large collections and portal-like resources.

The Humbul 'medieval' hits offer very specialized resources. Those illustrated above include an essay on 'Piers Plowman' and an Academy researching into medieval Spanish charters and parchments. The other search results include software to help with the study of medieval Latin paleography, an exhibition of medieval musical manuscripts, and scholarly articles in English and other languages. The Google resources again are much more general and broader, with links to gateway-type resources, information on medieval food, music, history, and picture collections.

The Humbul results for the 'poetry' search are again very specialized. Indeed the first two links on the list are not even in English, but in Galician-Portuguese and Spanish. Again Google offers a broader, more general view of poetry with anthologies, appreciation societies, and resources for working poets.

5.9 PSIgate http://www.psigate.ac.uk/

Psigate has resources in the physical sciences, including astronomy, chemistry, earth sciences, materials science, physics and the history and policy of science.

Results of the PSIgate searches

4 PSIgate results appear on the 1st page of Google results

3 PSIgate results appear on the 2nd page of Google results

2 PSIgate results appear on the 3rd page of Google results

16 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three PSIgate Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = polymers **PSIgate hits = 1,055 Google hits = 1,810,000**				
About plastics http://www.americanplasticscouncil.org/benefits/about_plastics/about_plastics.html	N	N	N	Y
Addition (Chain-growth) Polymers http://www.mpcfaculty.net/mark_bishop/addition_polymers.htm	N	N	N	Y
Advanced Engineering materials http://www3.interscience.wiley.com/cgi-bin/jhome/67500980	N	N	N	Y
Search term = spectroscopy **PSIgate hits = 1748 Google hits = 2,510,000**				
Acronyms and abbreviations [in Chemistry] http://www.chemie.fu-berlin.de/cgi-bin/acronym	N	N	N	Y
Advanced Chemical Experimentation and Instrumentation http://ocw.mit.edu/OcwWeb/Chemistry/5-33Advanced-Chemical-Experimentation-and-InstrumentationFall2002/CourseHome/index.htm	N	N	N	Y
Advanced Chemistry Development http://www.acdlabs.co.uk/	N	N	N	Y
Search term = nanotubes **PSIgate hits = 153 Google hits = 341,000**				
Carbon nanostructures : an efficient hydrogen storage medium for fuel cells http://www.fuelcelltoday.com/FuelCellToday/FCTFiles/FCTArticleFiles/Article_433_Carbon%20Nanostructures.pdf	N	N	N	Y
Carbon nanotube http://en.wikipedia.org/wiki/Carbon_nanotube	N	N	N	Y
Carbon nanotubes http://www.research.ibm.com/nanoscience/nanotubes.html	N	N	N	Y

Searching PSIgate

From the home page PSIgate offers a simple keyword search which can be limited to 'lecture notes and courses' or 'tutorials'. Further refinement is offered by an advanced search with resource type and subject options, or a range of other databases can be added to the resource base. Results are displayed alphabetically, and there is an extensive range of features.

Comments

Both the PSIgate and Google searches for 'polymers' produced quite specialized resources, often quite similar. Many of the PSIgate resources from the 'spectroscopy' search link to course materials from US University websites (the second one, for example, being an undated page from Monterey Peninsula College in California), whereas the Google hits offer more general information from commercial companies, government websites and university research centres. Both PSIgate and Google offer links to 'nanotube' research reports, articles, talks and descriptive websites from companies, universities and research organizations.

5.10 SOSIG http://www.sosig.ac.uk/

SOSIG covers the Social Sciences, including among other subjects, Economics, Education, European Studies, Geography, Politics, Law and Statistics

Results of the SOSIG searches

5 SOSIG results appear on the 1st page of Google results

1 SOSIG result appears on the 2nd page of Google results

19 results do not appear in the first three pages of Google results

All websites could be accessed from Google

First Three SOSIG Results	Found on Google page			Accessible from Google?
	1	2	3	
Search term = educational psychology **SOSIG hits = 94 Google hits = 2,880,000**				
Educational psychology interactive: readings in educational psychology http://chiron.valdosta.edu/whuitt/materials/elecfile.html	Y			
Leslie Owen Wilson's Curriculum, instruction and educational psychology home page http://www.uwsp.edu/education/lwilson/index.htm	N	N	N	Y
Mark Gover's Educational Psychology Homepage http://www.msu.edu/user/govermar/	N	N	N	Y
Search term = civil rights **SOSIG hits = 555 Google hits = 9,320,00**				
Harvard Civil Rights – Civil Liberties law review http://www.law.harvard.edu/studorgs/crcl_lawreview/	N	N	N	Y
Australian human rights and civil rights http://home.vicnet.net.au/~victorp/vphuman.htm	N	N	N	Y
Ko'aga Rone'eta: a journal of human rights http://www.derechos.org/koaga/main.htm	N	N	N	Y
Search term = poverty **SOSIG hits = 631 Google hits = 9,560,000**				
US poverty studies and poverty measurement: the past twenty-five years http://sticerd.lse.ac.uk/dps/case/CP/CASEpaper42.pdf	N	N	N	Y
PovertyNet http://worldbank.org/poverty/index.htm	Y			
Poverty net http://www.worldbank.org/poverty/	Y			

Searching SOSIG

The home page offers a simple keyword search box, and subject browsing. There is a link to an advanced search option which offers search refining by subject area and resource type, and search expansion by linking to the Social Science Search Engine. It is stated that results are ranked by the number of occurrences of the keyword in the record, but there is also a weighting or 'points' system which helps to push the most relevant results to the top of the list.

Comments

It is notable that the SOSIG results for the 'educational psychology' search contain numerous university departmental websites, and the websites of individual lecturers. The Google results are a mixed bag of educational websites, journals, careers advice and university departments.

The 'civil rights' search results in SOSIG offer a scholarly world wide view of civil rights issues, whereas the Google results tend to be more heavily weighted towards US government websites and US civil rights campaign issues.

SOSIG also offers more scholarly material and links from the 'poverty' search than the Google results, which again tend towards government sites and campaign groups. The first SOSIG resource is to a paper from an author at LSE: the link leads to a version of the paper dated 2000 which the author does not list in his personal bibliography. Instead he lists a version published in an academic journal in 2002. The second and third results from SOSIG point to an identical web site.

5.11 Searching using Resource Finder

An alternative to searching individual Hubs is to cross-search them all using the RDN Resource Finder. Although cross-searching was not initially planned to be an activity for this study, a small sample of searches was performed later at the request of the JISC. This produced interesting results. The first ten Resource Finder hits for a 'poetry' search, for example, are all from Humbul, but are quite different from those retrieved at the Hub. There is The Poetry Society, with resources 'pertaining to current poetry in Britain', or the Poetry Portal with resources for readers and writers.

The first nine Resource Finder hits for a 'migration' search come from SOSIG, and the tenth from GEsource. Two of the SOSIG resources are duplicated in the GEsource. This means, of course that a student restricting a search to a single Hub might be likely to miss a wealth of other possible relevant and useful resources hidden at the other Hubs (unless there is extensive duplication).

A search for 'diabetes' brings BIOME resources from the US National Institutes of Health, statistical information, the management of diabetes in pregnancy and nutritional advice.

A search for 'photography' finds resources from Artifact and SOSIG with many hits which are different from the Hub search.

The search strategy used with Resource Finder is clearly different from that used by some Hubs, and it does appear to rank resources with greater general appeal, and therefore perhaps more relevance to these simple searches, higher up the lists of hits. From these few searches it would certainly appear that it is worthwhile cross-searching the whole RDN in addition to carrying out searches at a single Hub.

Further investigation is needed to explore why all the differences noted occur. It is unclear at present whether the JISC or SPP has deliberately created tools which produce significantly different results in response to similar searches.

6. Conclusion and recommendations

This was always intended as a preliminary exercise, rather than a true comparative study of the RDN and Google. The key question asked was whether the trends in student searching behaviour which are well documented in the literature, namely the apparent preference for search engines as an information seeking tool, meant that students who use search engines are missing the academic resources of high quality which are available to them on the RDN Hubs.

None of the searches showed a high degree of correspondence between the Hub results and the Google results, so it might indeed be claimed that this is the case. However, this might well be too simplistic an interpretation. The fact that students do not find 'Hub resources' on Google does not necessarily mean that what they do find there is not of high quality or not of relevance to their academic work. Google results tend to rank highly websites from leading institutions such as governments, public bodies and large companies. The results are very broadly based and very plentiful. The Hubs often provide highly focused and specialized resources – a single journal article, a particularly interesting event, a learning resource targeted at a specific student group. They are full of the context which Google resources can lack, but being for such a specialized audience they may lack broad appeal, even within a particular subject area. The first two 'hits' of the Humbul 'poetry' search, for example, are not even in English; one is in Galician-Portugese, the second in Spanish, and there is no translation option. What would a student interested in 'poetry' make of this? The first hit of the BIOME 'museum' search is the winner of a wildlife photography competition. The Hubs offer plenty of opportunity to refine the search, but this infers a need for commitment on the part of the searcher and a degree of sophistication in searching behaviour which 'Google searchers' (i.e. most students!) seem to lack.

An alternative to searching individual Hubs is to cross-search them all using the RDN Resource Finder. This produces interesting results. The first ten hits for a 'poetry' search, for example, are all from Humbul, but are quite different from those retrieved at the Hub. There is The Poetry Society, with resources 'pertaining to current poetry in Britain', or the Poetry Portal with resources for readers and writers. A search for 'diabetes' brings BIOME resources from the US National Institutes of Health,

statistical information, the management of diabetes in pregnancy and nutritional advice. A search for 'photography' finds resources from Artifact and SOSIG with many hits which are different from the Hub search. The search strategy used with Resource Finder is clearly different from that used by some Hubs, and it does appear to rank resources with greater general appeal, and therefore perhaps more relevance to these simple searches, higher up the lists of hits.

Of course the Hubs are not like Google, nor was it intended that they should be. They do not fit easily with the 'get in, get stuff, get out' way of thinking. They are complex and offer academic context, and in order to ensure that the user gains the most benefit from them they require their users to sustain a certain amount of learning, familiarisation and commitment. For example, several of the Hubs rank their results alphabetically. It is clear from the study that this strategy produces some strange results. These may in themselves be of high quality but it is, for example, perplexing that the first result in a search for 'Manchester' is an article on an Australian participant in a 1977 cultural event, or for 'spectroscopy' is a list of acronyms and abbreviations. The only logical conclusion which can be drawn here, is that the most successful tactic for getting your website to rank highly in the Hub's list of hits is to give it a title which begins with a number, the letters 'aa', or a foreign script symbol. Again this might not matter if students were sophisticated searchers, but if they transfer their 'Google habits' to the Hubs, they are likely to quickly start searching elsewhere, or abandon this type of resource altogether.

It is also clear that some Hubs have a much larger resource base upon which to draw than others. Altis, of course, is quite new, but they and others may wish to consider making a positive effort to augment the hits for their most popular searches, in order to give users more choice and greater satisfaction.

We believe that this preliminary study has not answered our research question adequately, but it has raised some key questions, namely:-

- how do Hub users actually search the Hubs? Do they take time to learn how to use them, or do they apply previously learned behaviours to their searching? Do they use the many sophisticated searching options which the Hubs offer? Do they refine their searches or quickly go elsewhere? If they do go elsewhere, where do they go to?

- do users find useful resources, and if not do they return to the Hubs on future occasions?
- do regular users discover the Hubs for themselves, or are they introduced to them in a learning situation?
- to what extent are the resources presented by the hubs of high quality? Bearing in mind that quality always relates to context, would the results be judged by experts in the field to be 'the best' sources of information on the topics sought?

We believe that such questions merit further exploration and would recommend that further study be undertaken.

Bibliography

Banwell et. al. (2004) The JISC User Behaviour Monitoring and Evaluation Framework Journal of documentation vol 60 no 3

Brophy P, Fisher S, Griffiths J. R. and Markland M. (2001) EDNER: Formative Evaluation of the Distributed National Electronic Resource: DNER Service evaluation (Deliverable MDA 2, EDNER Project). Manchester: CERLIM (The Centre for Research in Library & Information Management). [Online] [cited 27 July 2004] http://www.cerlim.ac.uk/edner/dissem/da2a.doc

Brophy, P., Fisher, S., Jones, C.R. and Markland, M. (2004) *EDNER: Final Report.* Manchester: CERLIM (Centre for Research in Library & Information Management). [Online] [cited 27 July 2004] http://www.cerlim.ac.uk/edner/dissem/edner-final.doc

Google Technology homepage [online] [cited 27 July 2004] http://www.google.com/technology/

Griffiths, J and Brophy, P (2002) Student searching behaviour in the JISC Information Environment; Ariadne 33 available online at http://www.ariadne.ac.uk/issue33/edner/intro.html

Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the web. *SIGIR Forum*, Vol. 32. No. 1., pp. 5 -17.

Jones C., & Goodyear, P (2003). EDNER: Formative Evaluation of the Distributed National Electronic Resource: Pre-1992 University Institutional Case Study (Deliverable Z7-C-1, EDNER Project). Lancaster: CSALT (The Centre for Studies in Advance Learning Technologies) Lancaster University. http://www.cerlim.ac.uk/edner/dissem/c1.doc

OCLC (2002), 'White paper on the information habits of college students (research by Harris Interactive)', executive summary available at: http://www2.oclc.org/oclc/pdf/printondemand/informationhabits.pdf

RDN 'about RDN' [online] [cited 27 July 2004] http://www.rdn.ac.uk/about/

Spink, A et. al. (2002) From e-sex to e-commerce: web search changes; IEEE Computer p108 http://www.scils.rutgers.edu/~tefko/IEEEComputer2002.pdf